COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a new Movie Theatre in the city of Hyderabad

By: Roshan

January 2020



**Introduction**

Movies are a currently a great source of entertainment. It’s one of the best ways to come out of the modern stressful life. Considering the city of Hyderabad, there are many movie loving people and is also a great business. People love to watch all sorts of movies, from regional to worldwide from comedy to science fiction. Thus movie theatres are among the best businesses. There are many established theatres running successfully in the city. As the population in the city is on a rise, new theatres will have a good scope of development and can attract people. So one can find new theatres making use of this opportunity to earn money by meeting the demand. But a new one can’t simply be set up. Various factors may influence the setting up new theatres and one of the main factor is business competition. Lesser the competition more the success. The aim of this project is to find such a suitable location by considering the location of all the other theatres and finding a place where there are not many.

**Business Problem**

The objective of this capstone project to is to analyse and select the best location in the city of Hyderabad to open a new movie Theatre. Using data science methodology and machine learning techniques, the project aims at answering the question, what is the best place for a movie theatre businessman in the city of Hyderabad to open a new theatre, in terms of less business competition?

**Target audience of this project**

This project can be very useful for movie distributors or theatrical exhibitors who are looking to invest in or open a new movie theatre in the city of Hyderabad. As there are already a large number of theatres in the city it can be difficult to decide the right place to build a new one. There may be many other factors which can influence the building of a new one like population, incentives, etc… but one of the main factor is business competition, which directly effects the success of the theatre. Thus this project can help the movie distributors to filter out the areas having a low occupancy of theatres, and they can further think based on the listed out regions if necessary.

**Data Required**

Data required to the solve the problem is:

* Neighbourhood data of the city of Hyderabad, this includes all the place names. This data can be brought from the Wikipedia page [here](https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India).
* Latitude and longitude data of the list of places in the city. The python package Geocoder can be used for this.
* Last and the most important is the venue data of each place. This can be brought by using the Foursquare API.

Ways to extract the data:

* Firstly, for the neighbourhood data, the source is a Wikipedia page hence the data needs to be scraped, and the python beautifulsoup package and the requests package will be used for that purpose.
* The python Geocoder package can be of great use to get the latitude and longitude data of different places. Another python package, Folium is used for map illustrations.
* Venue details can be brought by making an API call to the foursquare data. For this an account is required with a client id and client secret, and the url for the call includes the same.
* The python package sklearn is used to implement machine learning algorithms in this case the KMean Clustering algorithm.